

March 6, 2025

***E-Filed***

The Honorable Susan van Keulen

United States District Court for the Northern District of California

San Jose Courthouse, Courtroom 6 – 4th Floor

280 South 1st Street

San Jose, CA 95113

Re: *In re Google Generative AI Copyright Litigation*; Master File Case No. 5:23-cv-03440-EKL; Consolidated Case No. 5:24-cv-02531-EKL

Dear Judge van Keulen:

The Parties submit this letter brief regarding disputed provisions in the proposed ESI Protocol (“ESIP,” Ex. A) and Protective Order (“PO,” Ex. B (clean) & Ex. C (changes tracked from model order)). Fact discovery closes in 225 days, on October 17, 2025. Counsel have met and conferred numerous times to narrow disputes. There remain 7 disputed provisions.

**Plaintiffs’ Position:** After consolidation, the parties began collective negotiations on these discovery orders. All Parties have made concessions and propose deviations from the model orders. Plaintiffs’ proposals on the few remaining disputes provide clarity, promote efficiency, and avoid conflict. Defendants’ proposals are designed to evade obligations, impose onerous restrictions, and limit Plaintiffs’ access to critical information. The Court should adopt Plaintiffs’ proposal.

**PO §2.17: Source Code.** Plaintiffs propose a precise definition of source code, excluding material not genuinely part of source code, like design documents, descriptive schematics, and comments. Plaintiffs’ definition, honed through meeting and conferring, incorporates elements specifically requested by Defendants, including pseudocode. Defendants’ proposal, though based on the Model, is overbroad and would give Defendants wide discretion to apply a source code designation to materials that lack the trade secret sensitivity of source code, restricting Plaintiffs’ ability to review and use such documents outside of the onerous protocols in §8, which far exceed the Model source code review procedures. These arduous procedures justify a precise definition. “Source Code” should be narrowly tailored to actual source code and not just any document Google deems to have technical significance. As Google notes, Plaintiffs’ counsel stipulated to the Model definition in prior AI litigation, but that led to over-designation and burdensome inspections, delaying meaningful discovery. This underlines the need for a precise and tailored definition. Plaintiffs’ proposal strikes an appropriate balance, sufficiently protecting truly sensitive source code and ensuring that such designations are not abused. Since the entry of the orders Google cites,

courts in this District have favored more precise language, tailored to the contours of complex AI cases. *See, e.g., In Re Mosaic LLM Litig.*, No. 24-cv-1451, Dkt. 62 (Dec. 5, 2024); *Bartz v. Anthropic*, No. 24-cv-5417, Dkt. 62 (Nov. 15, 2024); *Nazemian v. Dubus*, No. 24-cv-1454, Dkt. 99 (Jan. 8, 2025); *Concord Music Grp. v. Anthropic*, No. 24-cv-3811-EKL-SVK, Dkt. 293 (Jan. 27, 2025).

**PO §2.16: Export Control.** Export control information (“ECI”) is a statutorily-defined universe of ultra-sensitive material subject to strict regulatory control because its transfer or disclosure to foreign nationals or entities outside the U.S. is considered a potential threat to national security or foreign policy interests. *See* Export Admin. Reg., 15 C.F.R. §§ 730.1, *et seq.*; Int’l Traffic in Arms Reg., 22 C.F.R. §§ 120.1, *et seq.* Once material is classified as ECI, it cannot be handled by any non-U.S. citizen or taken outside the U.S. Individuals that violate ECI laws face civil and criminal penalties, including up to 20 years imprisonment and \$1 million in fines. *See*, 50 U.S.C.A. § 1705; 15 C.F.R. § 764.3. ECI restrictions would apply to the Court and litigants alike. Google insists that ECI restrictions be included, but cannot identify *any* relevant material that could feasibly be designated ECI, let alone any applicable statute, regulation, or export control classification number. Google’s handwaving at the recent Executive Order does not alter this. The case involves Google’s training of its AI products on copyrighted works like text and image art and literature, not state secrets. Courts in other AI copyright actions decline to impose ECI restrictions on this record. *See, e.g., Concord*, ECF 293; *Bartz*, ECF 62. The Court should do the same.

**ESIP: Preservation.** Defendants object to language reinforcing the Parties’ obligations to preserve relevant discovery. This is troubling, given Google’s prior preservation failures. *See In re Google Play Store Antitrust Litig.*, 664 F. Supp. 3d 981, 993-94 (N.D. Cal. 2023) (“Google intended to subvert the discovery process ... with the intent to prevent [information’s] use [and] to deprive

another party of the information's use in litigation." (quotations omitted)). Preservation is a foundational issue and should be included in the ESIP. *See* N.D. Cal ESI Guidelines § 2.01.

**ESIP: Privilege Logs.** The Parties must be able to evaluate privilege claims and address disputes before fact discovery closes. Plaintiffs' proposal requires substantial completion of document production and privilege logs 45 days before the close of discovery—ensuring that logs are provided with adequate time to resolve any disputes during discovery. Defendants' proposal—provide logs 45 days after production up to the close of discovery—significantly increases the risk of eleventh-hour privilege disputes, which are all-too common in cases against Google. *See, e.g., In Re Google RTB Consumer Priv. Litig.*, Dkt. 746 (appointing a Special Master to review and resolve disputes to Google's privilege logs after the close of fact discovery); *Oracle Am., Inc. v. Google Inc.*, 2011 WL 5024457, \*9 (N.D. Cal. Oct. 20, 2011) (affirming that Google improperly asserted privilege over key communications); *In re Google, Inc.* 462 Fed. Appx 975, 978-79 (Fed. Cir. 2012) (same). Plaintiffs' current request of a 45-day compliance deadline is not the "same" as what was previously proposed. ECF No. 84. It is necessary to ensure compliance with the case schedule and allows for meaningful review of privilege assertions. *See Loop AI Labs Inc. v. Gatti*, 2016 WL 2908415, at \*3 (N.D. Cal. May 13, 2016) (late privilege log deprived receiving party of ability to assess or challenge privilege claims).

There are two disputed metadata fields in § 10(d): File Name and Email Subject. These are auto-populated fields that provide information needed to assess privilege. *Ralls v. U.S.*, 52 F.3d 223, 225 (9th Cir. 1995) ("The party asserting [privilege] has the burden of establishing the relationship and the privileged nature of the communication."). These fields are called for by §7.d. of the Standing Order, which requires logs to contain the "title and description." In the unlikely event that privileged information is included in these fields, Plaintiffs' proposal permits exclusion.

**ESIP: Hyperlinks.** Hyperlinks are increasingly used over attachments. A custodian will likely have an email that says “use the version at [hyperlink].” If only the email is collected, and not the hyperlinked file, search terms will not hit the email, and this exchange will not be reviewed. Google can automatically identify and search the latest version of hyperlinked files stored on Google Drive, but says it cannot identify and search “point-in-time” versions. Plaintiffs’ proposal is tailored to these capabilities and limitations. It asks Google to apply search terms to latest versions of hyperlinked files, review and produce relevant documents, and provide metadata necessary to associate the hyperlinked document with its parent. And where an email hits on a search term, it asks Google to review any associated hyperlinked document. Plaintiffs understand that the current version may sometimes differ from the point-in-time version, but that does not mean it is no longer relevant. This is the automated, least burdensome method that Google has designed to retrieve hyperlinks. If a “point-in-time” version of the file is needed, Plaintiffs may ask Google to collect that version, which Google says it can do. Google raises only an unsubstantiated burden objection and cites cases in which the producing party lacks capability to produce hyperlinks. Plaintiffs refer the Court to *Google’s own representations about its capabilities*. Indeed, courts in this District have noted the propriety of using Google Vault to identify and produce hyperlinked files. *See, e.g., In re Uber Techs., Inc., Passenger Sexual Assault Litig.*, 2024 WL 3491760, at \*11 (N.D. Cal. Mar. 15, 2024). Plaintiffs’ proposal reflects Google’s capabilities and protects Plaintiffs’ access to relevant discovery.

**ESIP: Email Threading.** Plaintiffs agree that threading may be used for *review*. Plaintiffs dispute Google’s use of threading for *production*. Emails should be produced in the form they are normally kept, and from each custodian who possesses such emails. The length and number of participants in an email thread, like those often produced using thread suppression, often makes review by the

receiving party impracticable due to formatting and readability issues. If Google is permitted to use suppression for production, metadata from earlier, lesser-included emails will be omitted. Courts thus permit threading for review, but require production of all responsive emails, including earlier-in-time. *See, e.g., In re Insulin Pricing Litig.*, 2024 WL 2808083, at \*6 (D.N.J. May 28, 2024); *In re Actos Antitrust Litig.*, 340 F.R.D. 549, 552 (S.D.N.Y. 2022) (exclusion of lesser-included emails may result in loss of relevant information and metadata).<sup>1</sup>

**ESIP: Validation.** The Parties agree that Google will propose search terms and Plaintiffs will have input into such terms. As this process necessarily precedes the production of **most** documents, Plaintiffs are at a tremendous informational disadvantage. Plaintiffs will have no insight into, *e.g.*, acronyms or expressions that Google’s employees use, and limited ability to check whether the terms Google applies hit on a large portion of responsive material. Google’s position on validation is simply “trust us,” leaving Plaintiffs to rely solely on Defendants’ assessment of relevance and efficacy and creating real risk that key documents will never be reviewed. To ensure appropriate search terms are used, Plaintiffs’ proposed process emphasizes formal validation and quality control. Search term input *and* comprehensive validation are needed here to ensure that search methodologies effectively capture relevant materials while minimizing irrelevant results. *See* THE SEDONA PRINCIPLES, 3d Ed., cmt. 10.g (2018) (parties should develop “quality assurance to support the reasonableness of the review”); *In re Peanut Farmers Antitrust Litig.*, No. 19-cv-463, ECF Nos. 92 & 162 (E.D. Va.) (entering ESI protocol providing for search term input *and* a validation protocol); *In re Seroquel Prods. Liab. Litig.*, 244 F.R.D. 650, 652 (M.D. Fla. 2007) (noting that search terms process must be “cooperative” and that “[c]ommon sense dictates that sampling and

---

<sup>1</sup> If Google uses thread suppression at production, then where any thread member contains any additional or altered data not contained in the most-inclusive email (*e.g.*, attachments or BCC recipients), it is not a less-inclusive email and must be separately produced.

other quality assurance techniques must be employed.”). Plaintiffs do not seek to “micromanage,” only engage in a fair and collaborative process that ensures the capture of responsive materials.

**DEFENDANTS’ POSITION:** Defendants proposed a Protective Order and ESI Protocol in August and September 2024 that closely hewed to this District’s model orders, the Federal Rules, and precedents. Plaintiffs have demanded complicated and burdensome concessions, knowing that the costs of these additional demands will be borne disproportionately by Defendants. The Court should reject Plaintiffs’ overreaching add-ons and enter Defendants’ proposed orders.

**PO § 2.17: Source Code.** This district has “substantial expertise” in the law surrounding source code, “and its model order reflects the cumulative wisdom of the court and the bar.” *Telebuyer, LLC v. Amazon.com, Inc.*, 2014 WL 5804334, \*3 (W.D. Wash. July 7, 2014). Defendants’ source code definition mirrors the model order, encompassing both “computer code” and documents that “define or otherwise describe in detail the algorithms or structure of software or hardware designs,” where disclosure “would create a substantial risk of serious harm.” Ex. B § 2.17. Plaintiffs’ definition, limited to code “expressed in a formal programming language or through other structured notations,” *id.*, is unduly restrictive and inconsistent with precedent. *See Lone Star Silicon Innovations LLC v. Toshiba Corp.*, 2017 U.S. Dist. LEXIS 210484, \*4-10 (N.D. Cal. Dec. 20, 2017) (“[W]hat matters is not whether particular materials can accurately be described as literal ‘source code’ but whether they deserve protection *on par with* source code.”).

Courts in this district routinely adopt the model order’s definition, recognizing it as presumptively reasonable. *See, e.g., World Champ Tech LLC v. Peloton Interactive, Inc.*, 2022 WL 2159260, \*1 (N.D. Cal. Jun. 15, 2022); *Kelora Sys., LLC v. Target Corp.*, 2011 WL 6000759, \*2 (N.D. Cal. Aug. 29, 2011). They also reject attempts to narrow the model order’s definition absent specific, substantiated objections. *See In re Meta Pixel Healthcare Litig.*, 2023 WL 3483296, \*1

(N.D. Cal. May 16, 2023); *Lone Star*, 2017 U.S. Dist. LEXIS 210484, \*4-10 (rejecting argument “source code” should “cover only ‘computer code’ and materials ‘associated with computer code’”). Plaintiffs leverage the review protocols to which they agreed as a reason to narrow the definition, but they point to nothing in them that imposes any unreasonable burden.

Notably, Plaintiffs’ past practice and precedent support Defendants’ position. In other generative AI cases, plaintiffs—some represented by the same counsel here—have agreed to the model order’s source code provisions. *See In re OpenAI ChatGPT Litigation*, No. 3:23-cv-03223-AMO, Dkt. 106 ¶2.9 (N.D. Cal. Feb. 15, 2024); *Kadrey v. Meta Platforms, Inc.*, No. 3:23-cv-03417-VC, Dkt. 90 ¶2.9 (N.D. Cal. Jan. 25, 2024); *Doe I v. Github, Inc.*, No. 4:22-cv-06823-JST, Dkt. 63 ¶2.9 (N.D. Cal. March 7, 2023). And no example Plaintiffs cite, *supra* at 2, requires source code materials to be “expressed in a formal programming language,” as Plaintiffs would require. Plaintiffs’ concerns about convenience do not justify deviating from practice and precedent.

**PO § 12.6: Export Control Information.** Plaintiffs object to the inclusion of the model order’s export control provision, Ex. B § 12.6, arguing that no discovery “could feasibly be designated ECI.” *Supra* at 2. But the government has already restricted the export of certain generative AI-related information, and announced more to come. *See Framework for Artificial Intelligence Diffusion*, 90 Fed. Reg. 4544 (Jan. 15, 2025); Exec. Order No. 14179, 90 Fed. Reg. 8741 (Jan. 23, 2025). Given the risk of unauthorized export, and the severe penalties Google could face, *see* 50 U.S.C. § 4819(b)-(d), the model export control provision is needed.

**ESIP § 11: Search Term “Validation.”** The Parties have already agreed to a collaborative process to negotiate search terms from the outset. *See supra* at 5.<sup>2</sup> Plaintiffs seek to impose an

---

<sup>2</sup> For this reason, Plaintiffs’ sole, out-of-circuit judicial opinion is inapposite. *See In re Seroquel Prods. Liab. Litig.*, 244 F.R.D. 650, 662 (M.D. Fla. 2007) (defendants did *not* “work[] with Plaintiffs from the outset to reach agreement” but rather “undertook the task in secret”).



additional, burdensome layer of discovery on discovery, micromanaging Defendants’ production without any showing of deficiency. Ex. A § 11. Plaintiffs’ proposal demands Defendants undertake a multi-stage process involving: (1) reviewing a “statistically valid” “Null Set Sample”; (2) producing responsive documents from that sample; (3) developing new search terms; (4) running the original *and* new terms across the entire data set; (5) producing a hit report; and (6) potentially repeating this entire process, including additional “Null Set Sample” reviews and the application of even more search terms, at Plaintiffs’ discretion for “any ... justifiable reason.” *Id.* This process, potentially required multiple times, is highly problematic. As courts have repeatedly emphasized, “the *responding party* is the one best situated to decide how to search for and produce ESI responsive to discovery requests.” *In re Viagra Prod. Liab. Litig.*, 2016 WL 7336411, \*1 (N.D. Cal. Oct. 14, 2016) (emphasis added). Absent a specific showing of “insufficient or deficient” document production, *Jensen v. BMW of N. Am., LLC*, 328 F.R.D. 557, 566 (S.D. Cal. 2019)—a showing Plaintiffs have not made—“neither the requesting party nor the court should prescribe or detail the steps that a responding party must take.” Sedona Principles, 3d ed., 19 Sedona Conf. J. 1, Principle 123, 127, 131 (2018) (requiring “a showing of a specific deficiency” before mandating discovery processes). Plaintiffs’ pursuit of “perfection” in production is not required. *Alivecor, Inc. v. Apple, Inc.*, 2023 WL 2224431, \*2 (N.D. Cal. Feb. 23, 2023).

Defendants’ proposal requires each party to “take reasonable steps to validate its review process ... using quality control measures” to determine whether relevant ESI is missing, and, consistent with Defendants’ authorities, allows each side to request information about the other’s validation methods if they have “good cause to believe” responsive ESI is missing. Ex. A § 11.

**ESIP §§ 6(d), 10(g): Email Threading.** Plaintiffs demand production or logging of all lesser-included emails. Ex. A §§ 6(d) & 10(g). This demand would substantially expand the scope

of review and production, creating significant inefficiency, contrary to established practice. Defendants’ proposal—to produce only the “unique most inclusive copy and related family members,” *id.* § 6(d)—is widely accepted. *See, e.g., In re Social Media Adolescent Addiction/Personal Injury Prods. Liab. Litig.*, 2024 WL 1786293, \*12 (N.D. Cal. Feb. 20, 2024) (most inclusive emails).

**ESIP § 9: Hyperlinks.** Plaintiffs’ demand that Google collect “current version[s]” of internal hyperlinked documents and preserve associated metadata, Ex. A at § 9, is legally unsound and unnecessarily burdensome. Hyperlinked documents are not attachments. *See, e.g., In re Meta Pixel Healthcare Litig.*, 2023 WL 4361131, \*1 (N.D. Cal. June 2, 2023). Google is already “searching and producing [custodians’] Google Drive documents”; requiring it to collect such documents “through both a direct collection *and* a collection through hyperlinks ... would dramatically increase redundancies in the collection, increase costs, and delay discovery,” without any corresponding showing of relevance or proportionality. *Nichols v. Noom, Inc.*, 2021 WL 948646, \*4-5 (S.D.N.Y. 2021). Indeed, Plaintiffs’ demand would require Google to collect documents linked to in emails even before determining whether that email is responsive, which will result in collections of documents that are likely non-responsive. Further, there is no way to know whether a current version of a linked document matches the version that existed perhaps years earlier when the link was shared, or whether any version was ever accessed or viewed.

Defendants’ proposal—manual retrieval of an internal linked document upon specific, reasonable request, Ex. A at § 9—is practical and judicially endorsed, including by Plaintiffs’ own authority. *In re Uber Techs., Inc., Passenger Sexual Assault Litig.*, 2024 WL 1772832, \*4 (N.D. Cal. Apr. 23, 2024). This approach is “more robust than an automated system,” and limits the “reliability and validation” issues of automated retrieval. *In re Social Media Litig.*, 2024 WL

1786293, \*8; *see also In re Meta Pixel Litig.*, 2023 WL 4361131, \*1 (directing parties to “consider reasonable requests for production of hyperlinked documents on a case-by-case basis”).

**ESIP § 10(b): Interim Discovery Deadlines.** Plaintiffs’ interim deadlines by which document production and privilege logging shall be “substantially complete,” Ex. A § 10(b), are arbitrary. More importantly, Plaintiffs already sought similar arbitrary deadlines for “substantial completion of document productions and production of privilege logs” in their proposed scheduling order, ECF No. 84 at 4-5, which Judge Lee rejected in favor of a single fact discovery deadline. ECF No. 88 at 3. Plaintiffs should not be taking the same request to a different audience.

**ESIP § 10(d): Privilege Logs.** Plaintiffs’ demand for the inclusion of “File Name” and “Email Subject” metadata fields on privilege logs, unless those fields are themselves privileged, Ex. A § 10(d), imposes an undue burden. Privilege logs already must describe a document’s subject matter. Plaintiffs’ proposal not only requires a party to review and describe a document on a log, but also to scrutinize metadata to prevent disclosure of privileged content in those fields. That additional layer of review, not called for in the Court’s standing order, can significantly increase logging burden, and will disproportionately impact Defendants without increasing the fairness of the discovery process.

**ESIP § 5: Preservation Obligations.** Plaintiffs seek to insert into the ESI protocol a vaguely worded preservation obligation—that the parties have taken and will “take reasonable steps to preserve” documents as required by “applicable law,” Ex. A at § 5. That is an unnecessary departure from the model order. It offers no clarity on the scope of the Parties’ preservation obligations, and serves only to convert preservation disputes into ESI disputes.

Dated: March 6, 2025

Respectfully submitted,

By: /s/ Lesley E. Weaver

Lesley E. Weaver (State Bar No. 191305)

Anne K. Davis (State Bar No. 267909)

Joshua D. Samra (State Bar No. 313050)

**BLEICHMAR FONTI & AULD LLP**

1330 Broadway, Suite 630

Oakland, CA 94612

Telephone: (415) 445-4003

Email: lweaver@bfalaw.com

adavis@bfalaw.com

jsamra@bfalaw.com

Gregory S. Mullens (*pro hac vice*)

**BLEICHMAR FONTI & AULD LLP**

75 Virginia Road, 2<sup>nd</sup> Floor

White Plains, NY 10603

Telephone: (415) 445-4006

Email: gmullens@bfalaw.com

/s/ Joseph R. Saveri

Joseph R. Saveri (State Bar No. 130064)

Cadio Zirpoli (State Bar No. 179108)

Christopher K.L. Young (State Bar No. 318371)

Elissa A. Buchanan (State Bar No. 249996)

Evan Creutz (State Bar No. 349728)

**JOSEPH SAVERI LAW FIRM, LLP**

601 California Street, Suite 1505

San Francisco, California 94108

Telephone: (415) 500-6800

Facsimile: (415) 395-9940

Email: jsaveri@saverilawfirm.com

czirpoli@saverilawfirm.com

cyoung@saverilawfirm.com

eabuchanan@saverilawfirm.com

ecreutz@saverilawfirm.com

/s/ Ryan J. Clarkson

Ryan J. Clarkson (State Bar No. 257074)

Yana Hart (State Bar No. 306499)

Mark I. Richards (State Bar No. 321252)

Tiara Aveness (State Bar No. 343928)

**CLARKSON LAW FIRM, P.C.**

22525 Pacific Coast Highway

Malibu, CA 90265

Telephone: (213) 788-4050

Email: rclarkson@clarksonlawfirm.com

yhart@clarksonlawfirm.com

mrichards@clarksonlawfirm.com

tavaness@clarksonlawfirm.com

Tracey Cowan (State Bar No. 250053)

**CLARKSON LAW FIRM, P.C.**

95 Third Street, Second Floor

San Francisco, CA 94103

Telephone (213) 788-4050

Email: tcowan@clarksonlawfirm.com

/s/ Stephen J. Teti

Stephen J. Teti (*pro hac vice*)

**LOCKRIDGE GRINDAL NAUEN PLLP**

265 Franklin St., Suite 1702

Boston, MA 02110

Tel.: (617) 456-7701

Email: sjteti@locklaw.com

Brian D. Clark (*pro hac vice*)

Laura M. Matson (*pro hac vice*)

Arielle Wagner (*pro hac vice*)

Eura Chang (*pro hac vice*)

**LOCKRIDGE GRINDAL NAUEN PLLP**

100 Washington Avenue South, Suite 2200

Minneapolis, MN 55401

Telephone: 612)339-6900

Facsimile: (612)339-0981

Email: bdclark@locklaw.com

lmmatson@locklaw.com

aswagner@locklaw.com

echang@locklaw.com

Matthew Butterick (State Bar No. 250953)

**BUTTERICK LAW**

1920 Hillhurst Avenue, #406

Los Angeles, CA 90027

Telephone: (323) 968-2632

Facsimile: (415) 395-9940

Email: mb@buttericklaw.com

*Counsel for Individual and Representative  
Plaintiffs and the Proposed Class*

Dated: March 6, 2025

/s/ David H. Kramer

David H. Kramer, SBN 168452

Email: dkramer@wsgr.com

Maura L. Rees, SBN 191698

Email: mrees@wsgr.com

**WILSON SONSINI GOODRICH & ROSATI**

**Professional Corporation**

650 Page Mill Road

Palo Alto, CA 94304-1050

Telephone: (650) 493-9300

Eric P. Tuttle, SBN 248440

Email: eric.tuttle@wsgr.com

**WILSON SONSINI GOODRICH & ROSATI**

**Professional Corporation**

701 Fifth Avenue, Suite 5100

Seattle, WA 98104-7036

Telephone: (206) 883-2500

*Counsel for Defendants*

GOOGLE LLC AND ALPHABET INC.

**SIGNATURE ATTESTATION**

I, Lesley E. Weaver, am the ECF User whose ID and password are being used to file this document. In compliance with N.D. Cal. Civil L.R. 5-1(i)(3), I hereby attest that the concurrence in the filing of this document has been obtained from the other signatory.

By: /s/ Lesley E. Weaver  
Lesley E. Weaver